DETAILED ACTION

1.   This action is responsive to the following communications: BPAI Decision,

issued 03/08/2010.

2.   Claims 1-8, 10, 12-21, 23-26, and 28-30 are allowed. Claims 1, 10, 15, 21, and

26 are independent claims.

EXAMINER'S AMENDMENT

An examiner's amendment to the record appears below. Should the changes

and/or additions be unacceptable to applicant, an amendment may be filed as provided

by 37 CFR 1.312. To ensure consideration of such an amendment, it MUST be

submitted no later than the payment of the issue fee.

Authorization for this examiner's amendment was given in a telephone interview

with Viktor Simkovic on 06/02/2010.

Please amend the claims as follows:

1.   (Currently amended) A method for crawling documents, performed by one

or more server devices, the method comprising:

receiving, by one or more processors associated with the one or more server

devices, a uniform resource locator (URL);

receiving, by one or more processors associated with the one or more server

devices, at least two different copies of a document associated with the URL; and

determining, by one or more processors associated with the one or more server
devices, whether a web site corresponding to the URL uses session identifiers based on
a comparison of URLs that are within the document and that change between the at
least two different copies of the document, where the web site is determined to use
session identifiers when a portion of the URLs that change between the at least two
different copies of the document is greater than a threshold.

2.   (Original) The method of claim 1, wherein the document is a home page of
the web site.

3.   (Previously presented) The method of claim 1, further comprising:
        extracting, when the URL corresponds to a web site that uses session identifiers,
a session identifier from the URL to obtain a clean URL; and
        determining whether the URL has already been crawled based on a comparison
of the clean URL to a set of clean URLs that represent previously crawled URLs.

4.   (Previously presented) The method of claim 3, wherein the compared
URLs that change include URLs that are local to the web site.

5.   (Original) The method of claim 3, wherein the session identifiers from the
URLs are extracted using rules for the web site.

6.   (Original) The method of claim 5, wherein the rules are determined automatically.

7.   (Original) The method of claim 3, further comprising:

receiving the URL as a URL from a previously crawled web document.

8.   (Original) The method of claim 3, further comprising:

crawling the URL when the URL is determined to not already have been crawled.

9.   (canceled)

10.   (Currently amended) A method for identifying web sites that use session identifiers, performed by one or more server devices, the method comprising:

downloading, by one or more processors associated with the one or more server devices, at least two different copies of at least one document from a web site;

extracting, by one or more processors associated with the one or more server devices, uniform resource locators (URLs) from the two different copies of the web document;

comparing, by one or more processors associated with the one or more server devices, the extracted URLs of the two different copies of the document; and

determining, by one or more processors associated with the one or more server devices, whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

11.   (canceled)

12.   (Original) The method of claim 10, wherein extracting URLs from the two different copies of the document includes extracting only URLs that are local to the web site.

13.   (Original) The method of claim 10, wherein the document is a home page of the web site.

14.   (Original) The method of claim 10, further comprising:

analyzing the extracted URLs, when the web site is determined to use session identifiers, to generate at least one rule identifying how the session identifiers are embedded in the URLs.

15.   (Currently amended) A device comprising:

a memory to store instructions; and

a processor to execute the instructions to implement:

a spider component configured to crawl web documents associated with at least one web site; and

a session identifier component configured to determine whether the web site uses session identifiers based on a comparison of a portion of uniform resource locators (URLs) that change between different copies of at least one web document downloaded from the web site.

16.   (Original) The device of claim 15, wherein the spider component further comprises:

at least one fetch component configured to download content from a network; and

a content manager configured to extract URLs from the downloaded content.

17.   (Original) The device of claim 16, wherein the spider component further comprises:

a URL manager configured to store the extracted URLs.

18.   (Original) The device of claim 15, wherein the at least one web document is a home page of the web site.

19.   (Original) The device of claim 15, wherein the portion of the URLs that change are identified from URLs that are local to the web site.

20.   (Original) The device of claim 15, further comprising:

a session rule generator configured to generate rules describing how the web

site embeds session identifiers in the at least one web document.

21.   (Previously presented) A device comprising:

means for downloading at least two different copies of at least one web

document from a web site;

means for extracting uniform resource locators (URLs) from the two different

copies of the web document;

means for comparing the extracted URLs of the two different copies of the web

document; and

means for determining whether the web site uses session identifiers when the

comparison indicates that at least a portion of the URLs change between the two

different copies.

22.   (canceled)

23.   (Original) The device of claim 21, wherein the means for extracting URLs

from the two different copies of the web document includes means for extracting only

URLs that are local to the web site.

24.   (Original) The device of claim 21, wherein the web document is a home page of the web site.

25.   (Original) The device of claim 21, further comprising:

means for analyzing the extracted URLs, when the web site is determined to use session identifiers, to generate rules describing how the session identifiers are embedded in the URLs.

26.   (Currently amended) A computer-readable medium One or more memory devices containing programming instructions that, when executed by at least one processor cause the processor to perform a method for identifying web sites that use session identifiers, the one or more memory devices including:

one or more instructions to download downloading at least two different copies of at least one document from a web site;

one or more instructions to extract extracting uniform resource locators (URLs) from the two different copies of the document;

one or more instructions to compare comparing the extracted URLs of the two different copies of the web document; and

one or more instructions to determine determining whether the web site uses session identifiers when the comparison indicates that at least a portion of the URLs change between the two different copies.

27.   (canceled)

28.   (Currently amended) The ~~computer-readable medium~~ one or more memory devices of claim 26, wherein the one or more instructions to extract ~~extracting~~ URLs from the two different copies of the web document includes one or more instructions to extract ~~extracting~~ only URLs that are local to the web site.

29.   (Currently amended) The ~~computer-readable medium~~ one or more memory devices of claim 26, wherein the web document is a home page of the web site.

30.   (Currently amended) The ~~computer-readable medium~~ one or more memory devices of claim 26, further comprising ~~instructions that cause the at least one processor to~~:
     one or more instructions to analyze the extracted URLs, when the web site is determined to use session identifiers, to generate at least one rule describing how the session identifiers are embedded in the URLs.

31.   (canceled)

REASONS FOR ALLOWANCE

The following is an examiner's statement of reasons for allowance:

Claims 1-8, 10, 12-21, 23-26, and 28-30 are allowed pursuant to the BPAI
Decision, issued 03/08/2010.

The above examiner's amendments to the claims are for the purpose of
rendering the claims statutory under 35 U.S.C. 101.

Any comments considered necessary by applicant must be submitted no later
than the payment of the issue fee and, to avoid processing delays, should preferably
accompany the issue fee. Such submissions should be clearly labeled "Comments on
Statement of Reasons for Allowance."

Conclusion

Any inquiry concerning this communication or earlier communications from the
examiner should be directed to AMELIA RUTLEDGE whose telephone number is
(571)272-7508. The examiner can normally be reached on Monday - Friday 9:30 - 6:00.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's
supervisor, Doug Hutton can be reached on 571-272-4137. The fax phone number for
the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see http://pair-direct.uspto.gov. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a

USPTO Customer Service Representative or access to the automated information

system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/Amelia Rutledge/
Primary Examiner, Art Unit 2176